Analysis of Subgoal Data in Computer Science Principles - Data Cleaning

Hari Ramilison, Graduate Student, Management Information Systems
Faculty Mentor: Briana Morrison, Computer Science

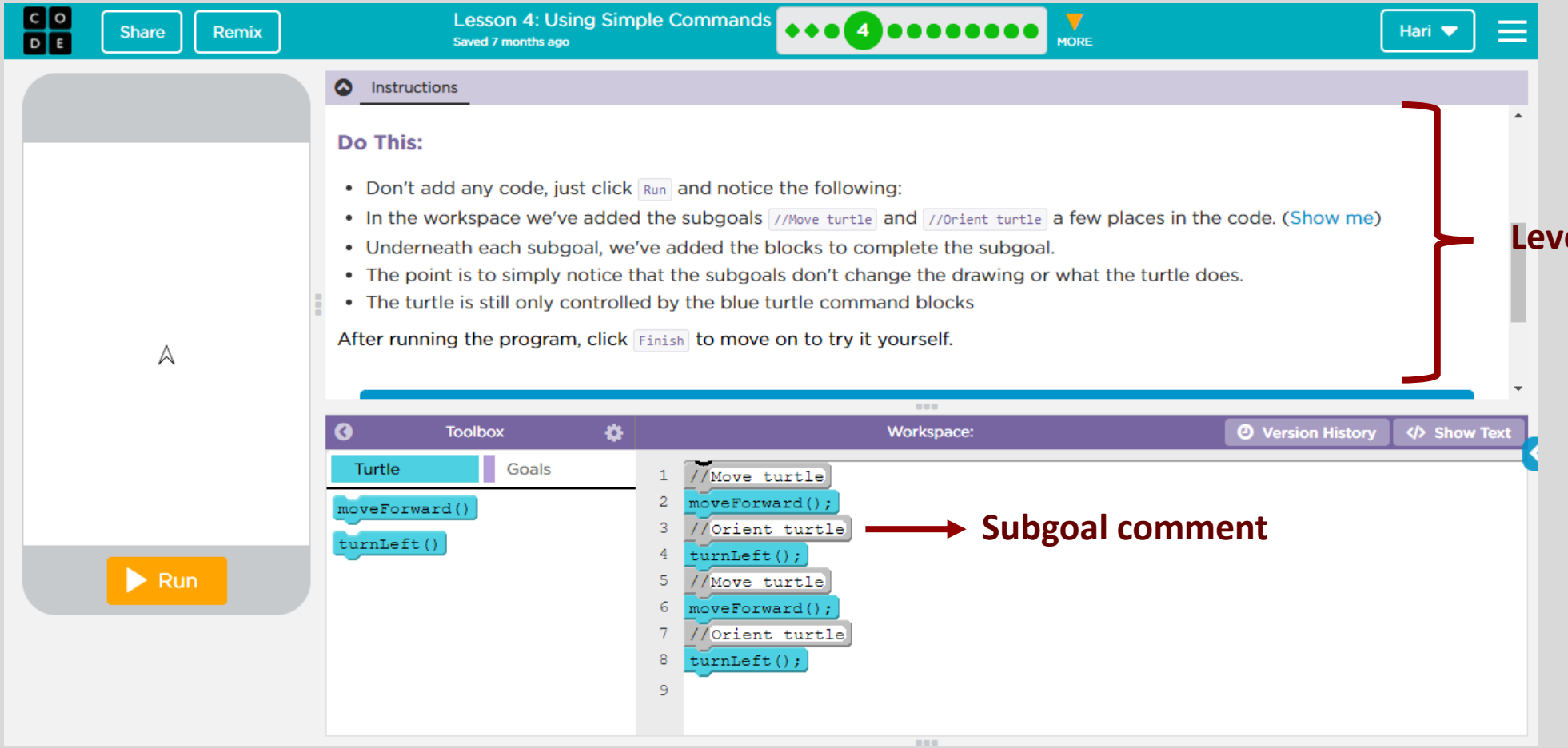
- Introduction
- Data Extraction
- Data Transformation
- Data Cleaning
- Data Validation
- Conclusion

• Introduction

- Evaluation of the effectiveness of students learning to solve programming problems.
- Use of subgoal labels within Code.org Computer Science Principles online course.

- Creation of automated processes to transform and clean the data from Code.org.

- # Data extraction



Reference: Code.org - Lesson 4: Using Simple Commands - Introducing Subgoals

• Data extraction

- Understanding data structure
  - Nine tables
  - Three types of assessments
    - App Lab levels
    - Multiple-Choice questions
    - Free response levels
  - Different types of subgoal comments
- Challenge
  - Scattered data are difficult to analyze

Data transformation

- Convert the data structure to two-dimensional table for each assessment type.

	AppLab	Free Response	Multiple-choice
Level	10	21	79
Subgoal comment or keyword	9	10	
Number of columns	90	210	79

• Data transformation

- Automated processes developed in SQL stored procedures
  - Create two-dimensional tables
    - Iterate on the list of levels
    - Iterate on the list of subgoal comments
    - Create a column for each level with each subgoal comment
      - E.g.: Level1_loop, Level1_moveturtle
  - Migrate data
    - Iterate on the list of students, levels and responses
    - Insert each value into its corresponding cell

• Data cleaning

- Remove unnecessary data
  - Data for students who did not answer
    - Five or more applab levels
    - Four or more multi-choice questions
    - Four or more free response levels
- Remove inconsistency
  - Data for students who were not found in all three tables

• Data validation

- Check the number of columns
- Randomly check the accuracy of the cells

• Research results

- How do students use subgoal comment blocks throughout the unit?

Subgoal Prompted	# times subgoal used (average # of students)	% of students
Move turtle (M = 11.4, SD = 13.6)	1-3 times (50 students)	8%
	4 times (527)	26%
	5-16 times (30)	18%
	17-20 times (100)	19%
	21-100 times (<18)	15%
Orient turtle (M = 10.3, SD = 12.8)	1-2 times (45)	4%
	3 times (488)	24%
	4-14 times (36)	19%
	15-19 times (92)	21%
	20-84 times (<20)	15%
Define function (M = 1.4, SD = 2.6)	1-2 times (200)	19%
	3-8 times (50)	14%
	9-16 times (10)	5%
Call function (M = 1.5, SD = 4.1)	1-2 times (200)	20%
	3-8 times (25)	7%
	9-45 times (<10)	5%

Subgoal Explained	# times subgoal used (average # of students)	% of students
Move turtle (M = 1.4, SD = 2.6)	1-5 times (50 students)	21%
	6-22 times (<14)	9%
Orient turtle (M = 1.7, SD = 2.1)	1 time (113)	6%
	2-4 times (75)	11%
	5-10 times (<20)	4%
Define function (M = 0.4, SD = 0.8)	1 time (157)	8%
	2 times (38)	2%
	3-9 times (<10)	1%
Call function (M = 0.4, SD = 0.8)	1 time (156)	8%
	2-9 times (<12)	3%

Subgoal None	# times subgoal used (average # of students)	% of students
Move turtle (M = 1.2, SD = 4.9)	1-4 times (50 students)	10%
	5-47 times (<15)	7%
Orient turtle (M = 1.0, SD = 4.4)	1-2 times (65)	7%
	3-8 times (15)	6%
	9-65 times (<10)	3%
Define function (M = 1.0, SD = 3.3)	1 time (121)	6%
	2-8 times (30)	11%
	9-45 times (<10)	3%
Call function (M = 1.0, SD = 3.4)	1-2 times (150)	15%
	3-7 times (40)	9%
	8-25 times (<10)	3%

Reference: Subgoal Redesign of Code.org's Intro to Programming Unit in Computer Science Principles

Thank you!